Yandex Cloud

Public Cloud — гайд по масштабированию

Нарек Татевосян, Cloud Solutions Architect, Yandex.Cloud





Обо мне



- Эксперт в инфраструктуре, сетях и контейнерах CKA, CCNP Enterprise, GCP Architect, RHCE
- Построил облако в Казахстане (Транстелеком) и развиваю облачную платформу в Yandex.Cloud
- Участвовал в проектировании и внедрении 100+ проектов для Enterprise и Service Provider, живущих до сих пор в продакшне

Кому будет полезен этот доклад



ИТ-специалисту, менеджеру и архитектору, который:

- жил в мире on-prem, хостингов и Vmware-based облаков
- столкнулся с миром публичных облаков

Содержание

Что такое публичное облако в 2021 году

С точки зрения Gartner

Magic Quadrant
for Cloud Infrastructure
and Platform Services



Challengers

Leaders

Amazon Web Services

Microsoft

Google

Alibaba Cloud

Oracle

Tencent Cloud

IBM

Niche Players

Visionaries

Ability to Execute

Completeness of Vision

As of August 2020

Источник gartner.com

Основные покупатели — это web-сервисы

Top Cloud Hosting Buyers*

Most of the top global cloud hosting buyers are concentrated in media and internet services.
We highlighted nine of the largest spenders by their estimated monthly spend on cloud hosting services.



$59M / month



$5M / month



$24M / month



$6M / month



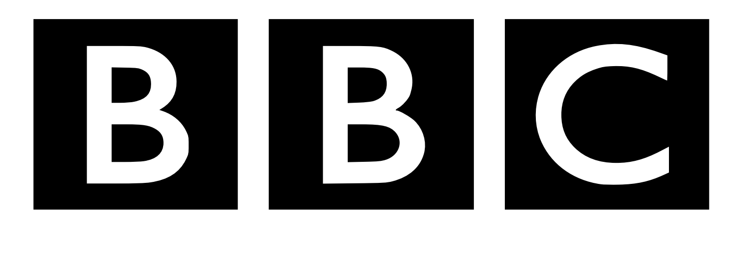
$20M / month



$3M / month



$16M / month



$6M / month



$3M / month

*Intricacy report content.intricately.com/2019-cloud-market-share-report

Картинка в Yandex.Cloud

Top-50 клиентов

- 50% клиентов — это SaaS
- 50% клиентов — это enterprise applications

Workloads

- ~40% крупных клиентов используют k8s (минимум non-prod сред)
- Но только ~25% ресурсов compute - это k8s, т. е. многие сидят еще на ВМ или ВМ и контейнерах

Почему вообще облако



Lower Time to Market

Automate everything

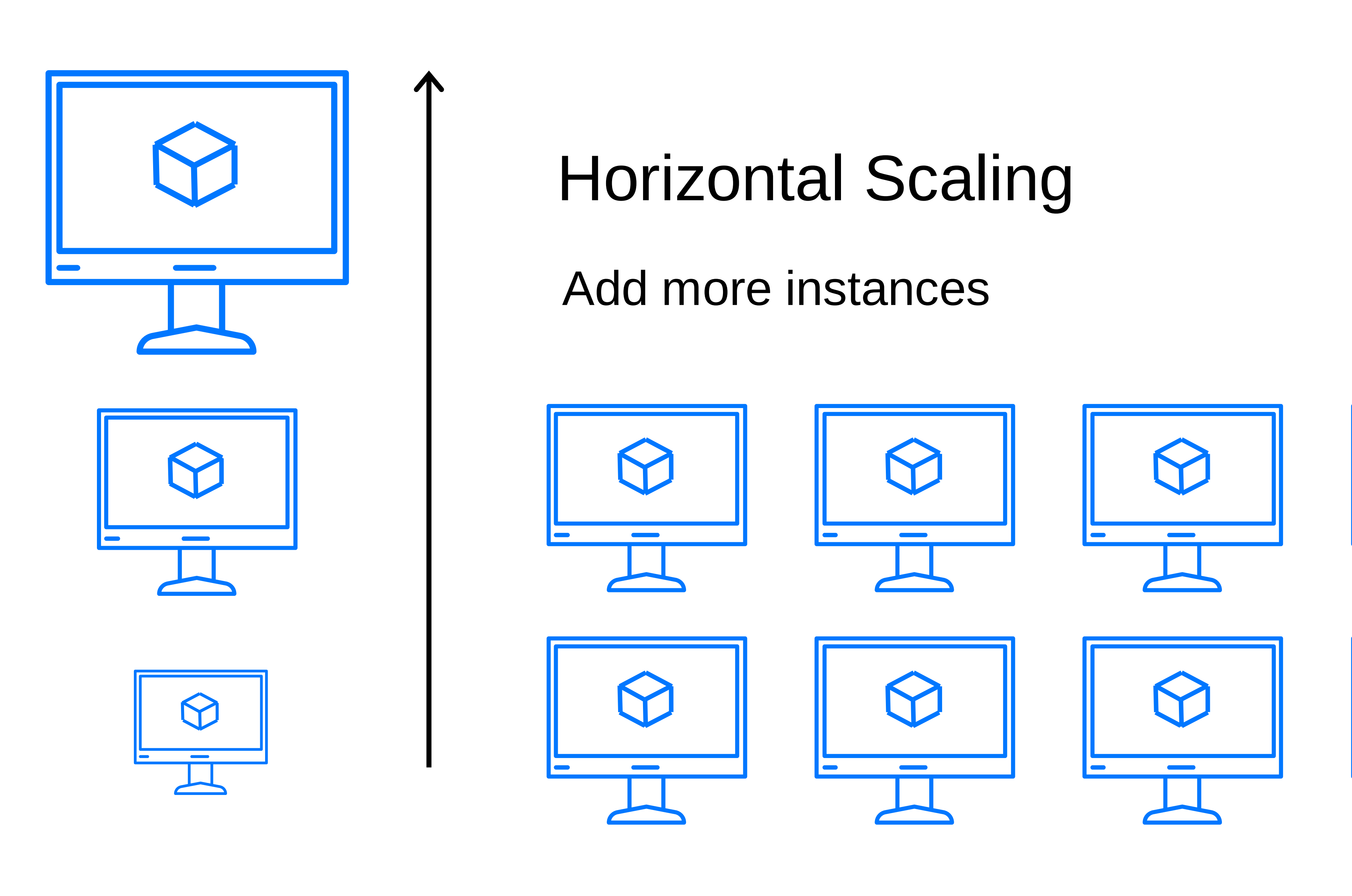
Scale everything

Изменение парадигм в облаке

Horizonal Scaling vs Vertical Scaling

Рекомендации:

- Разделите свои workloads на stateless и stateful
- Stateless workloads автомасштабируются горизонтально
- Stateful workloads как минимум вручную масштабируются вертикально

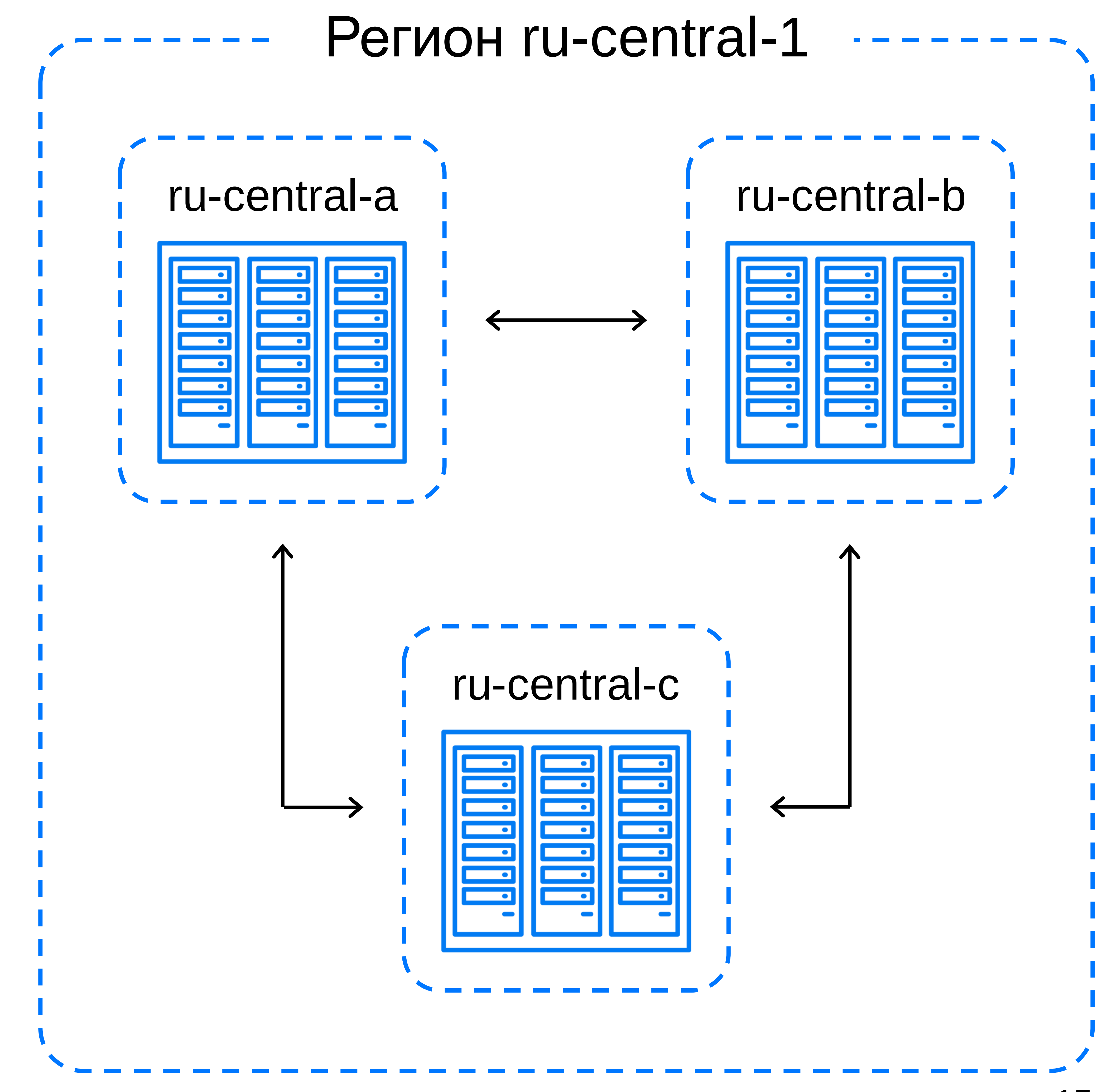


Зоны доступности и регионы вместо стоек и ЦОД

1. Зона доступности — обычно ДЦ
2. Регион — объединение ДЦ по географическому принципу

Рекомендации

› Изучите scope сервисов, которые вы планируете использовать: есть сервисы зональные, есть региональные, а есть глобальные (а еще у сервиса могут быть разные режимы работы)



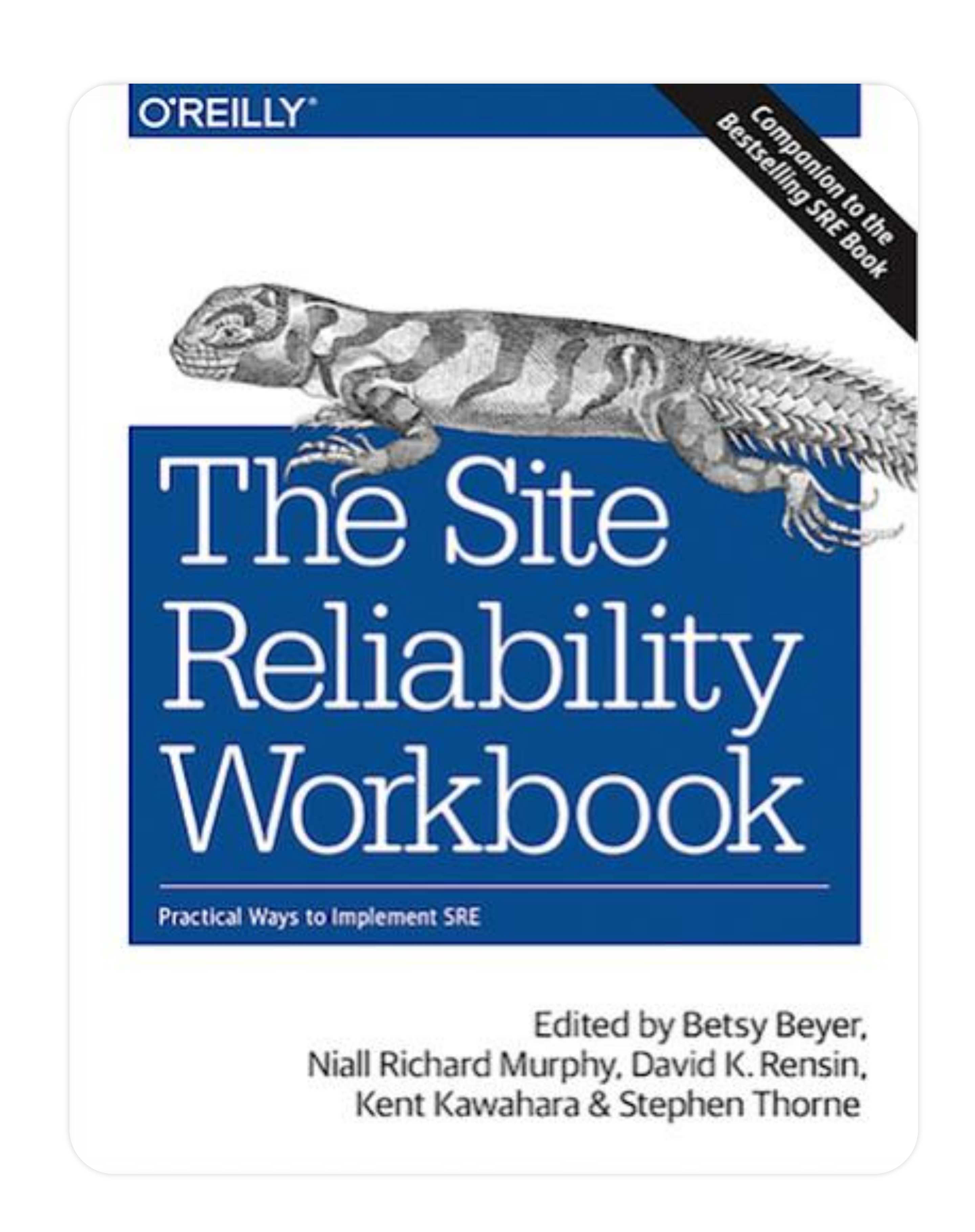
Scaling! = High Availability

Scale и облака прибавляют новых сценариев отказа

1. Каскадные сбои
2. DDoS от мониторинга
3. Апдейты в зоне доступности

Рекомендация

› Закладывайте сценарии отказа в сайзинг проекта





sre.google/books/

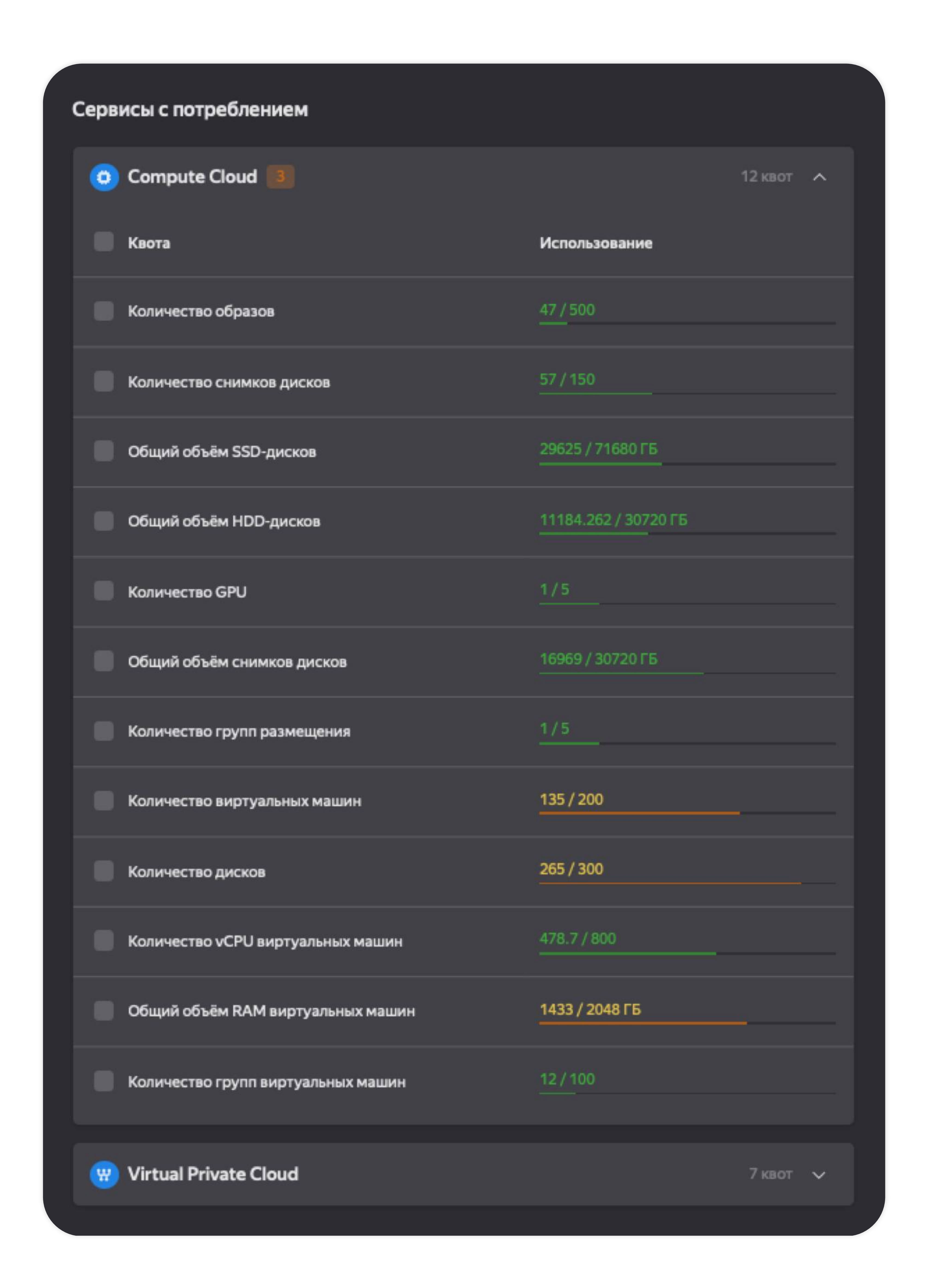
Квоты

Compute Engine quotas

You can use a Compute Engine resource up to its quota. Google Cloud Platform projects have separate Compute Engine quotas. If you reach a resource quota, you can request an increase to use more of that resource. Learn more

Request increase

Resource	Percent used		Use
Networks		100%	5 of 5
In-use IP addresses **global**		87%	20 of 23
Static IP addresses **us-central1**		86%	6 of 7
Target pools		82%	41 of 50
Instance templates		81%	81 of 100
In-use IP addresses **europe-west1**		78%	18 of 23
Managed instance groups **us-central1**		78%	39 of 50
Health checks		76%	38 of 50
CPUs **europe-west1**		71%	17 of 24
Total Local SSD disk reserved (GB) **us-central1**		70%	7,125 of 10,240
Backend services		62%	31 of 50
Forwarding rules		60%	30 of 50
CPUs **asia-east1**		58%	14 of 24
Firewall rules		55%	55 of 100
In-use IP addresses **asia-east1**		48%	11 of 23
VPN tunnels		40%	4 of 10
Instance groups **us-central1**		40%	60 of 150
Routes		36%	36 of 100
URL maps		36%	18 of 50
Total SSD disk reserved (GB) **us-central1**		35%	718 of 2,048
Target HTTP proxies		32%	16 of 50
SSL_CERTIFICATES		30%	15 of 50
Static IP addresses **global**		29%	2 of 7



Ресурсы не ограничены, но есть квоты

Что это

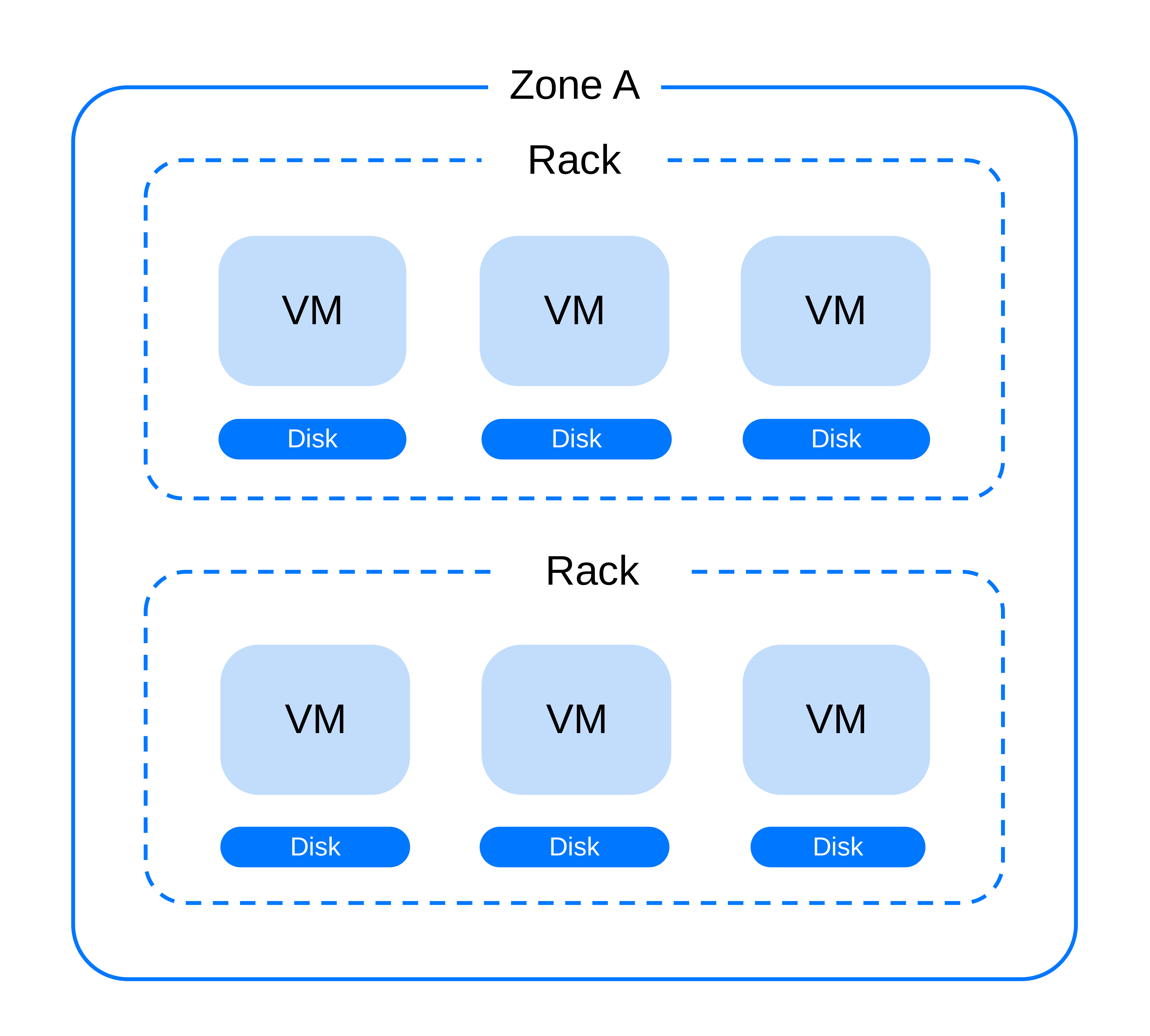
- Квота — объем ресурсов, которые можно использовать. Квота обычно повышается через саппорт (то есть небыстро)

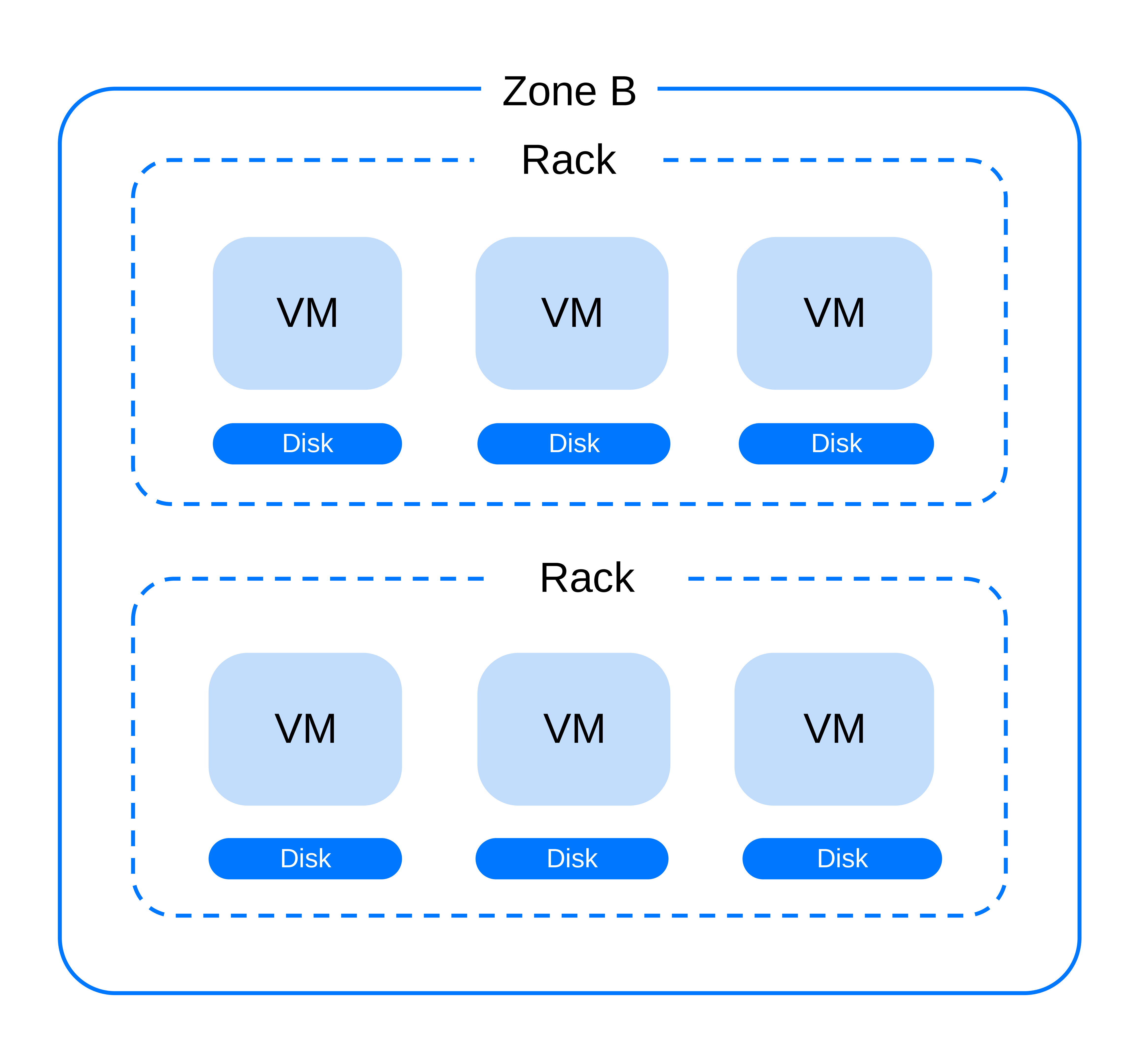
Рекомендации

- Внимательно читайте раздел о квотах и лимитах всех сервисов, которые планируется использовать
- Постоянно следите за квотами, чтобы избежать инцидентов

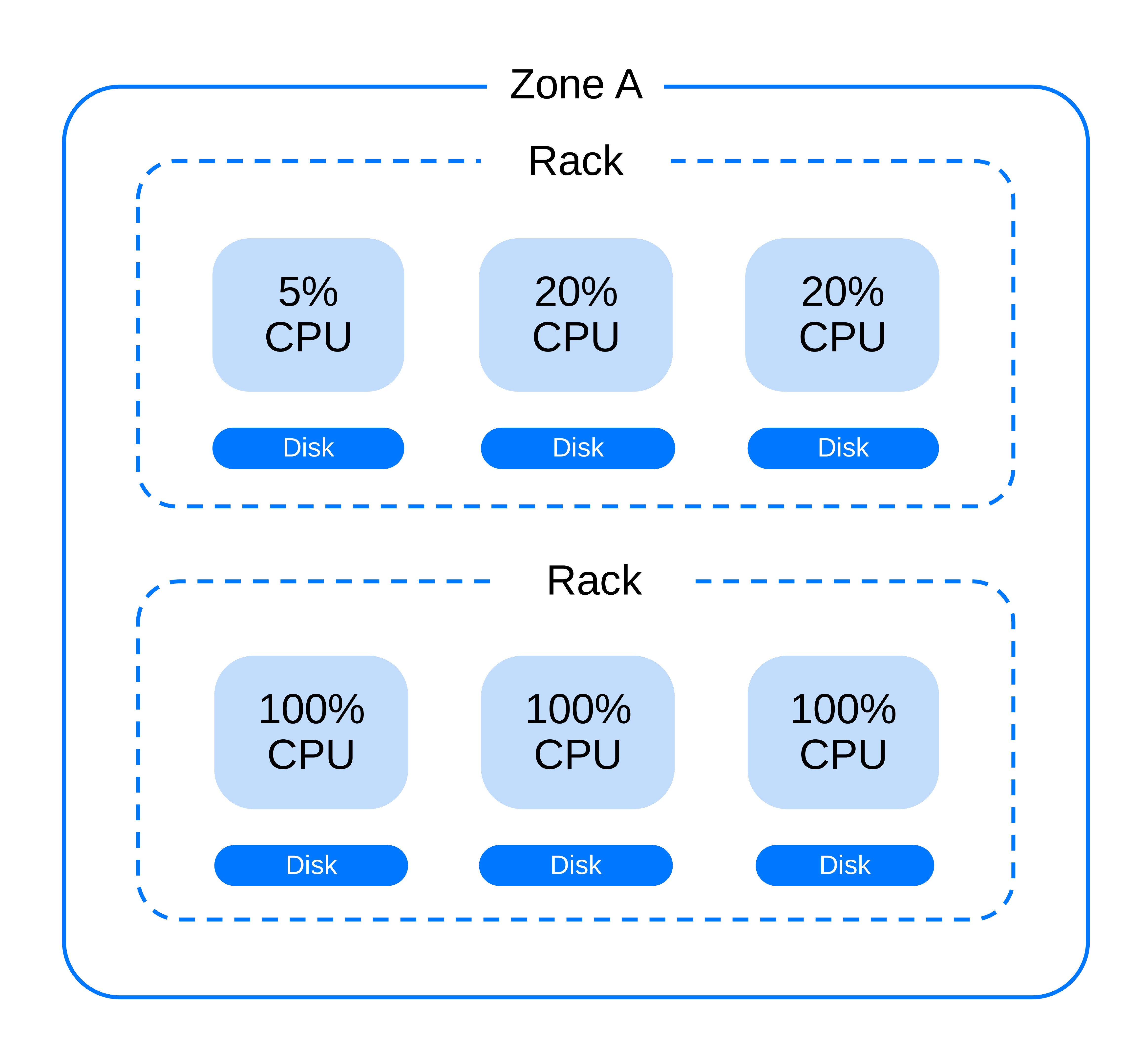
Чек-лист про IaaS

Compute VM


Zone A
Rack
VM
VM
VM
Disk
Disk
Disk
Rack
VM
VM
VM
Disk
Disk
Disk

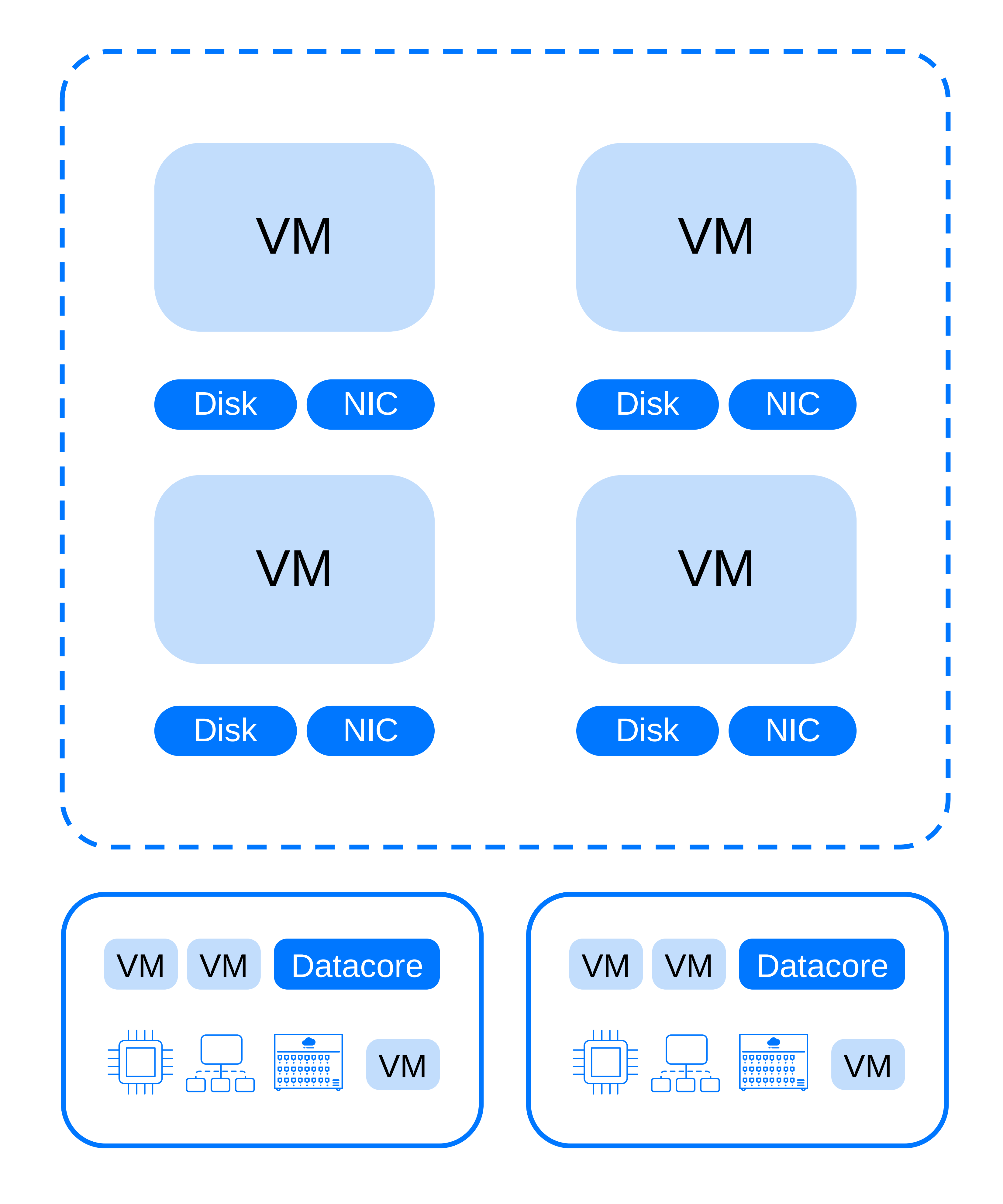

Zone B
Rack
VM
VM
VM
Disk
Disk
Disk
Rack
VM
VM
VM
Disk
Disk
Disk

Compute VM: переподписка

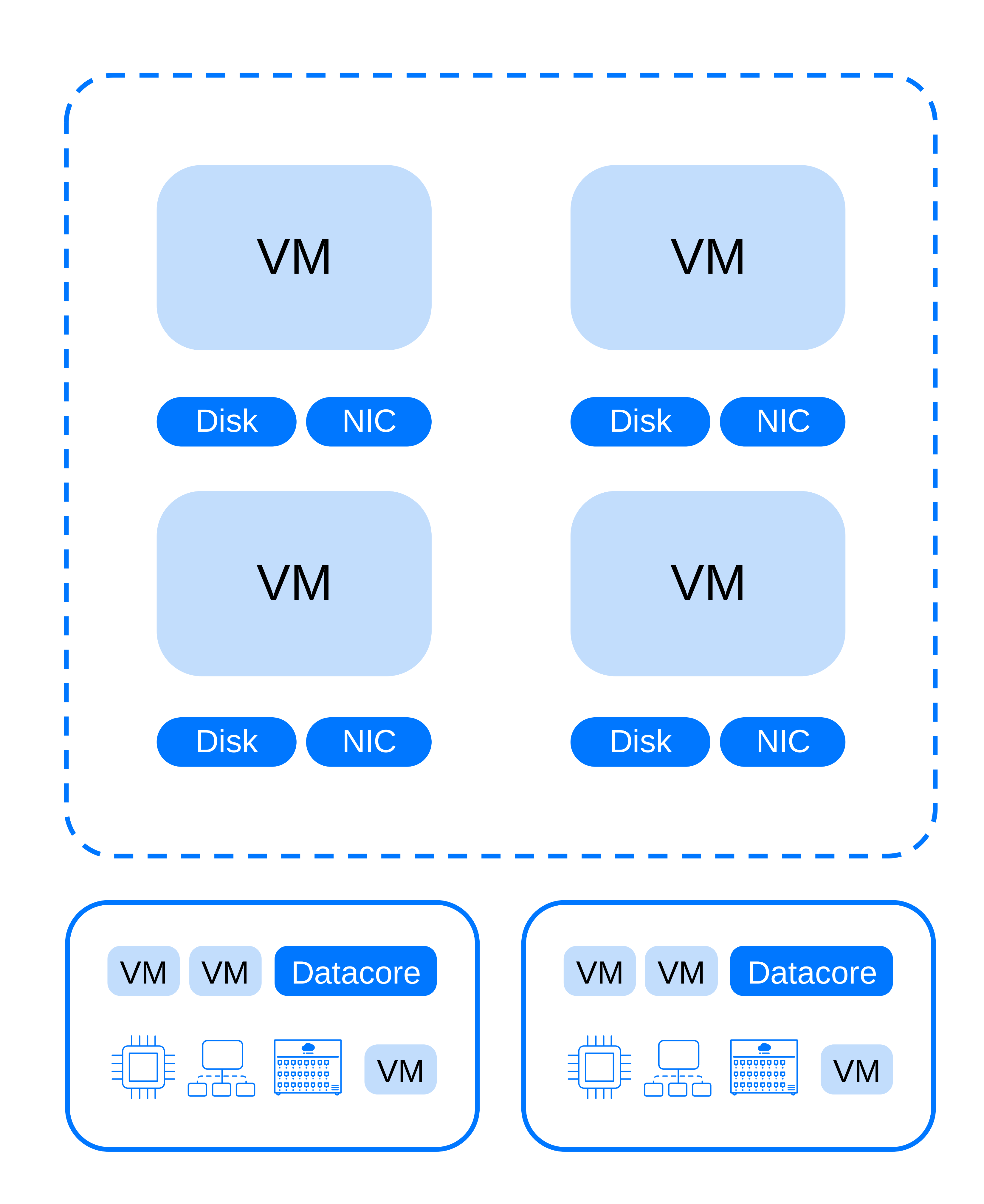


Compute VM — платформы

General purpose platform Intel Broadwell



GPU platform Intel Cascade Lake



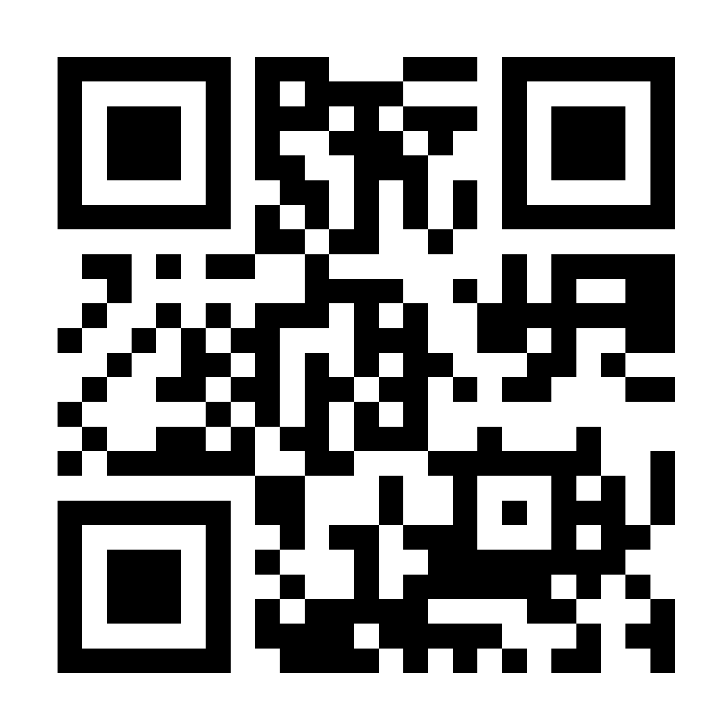
Compute VM

Особенности

- Обычно Зональные ВМ с x86 или ARM-архитектурой не могут мигрировать в другой ДЦ
- Существуют режимы с гарантированным и негарантированным резервированием CPU
- Отличаются платформами (вендор/модель CPU, число CPU на хост, соотношение CPU/RAM)

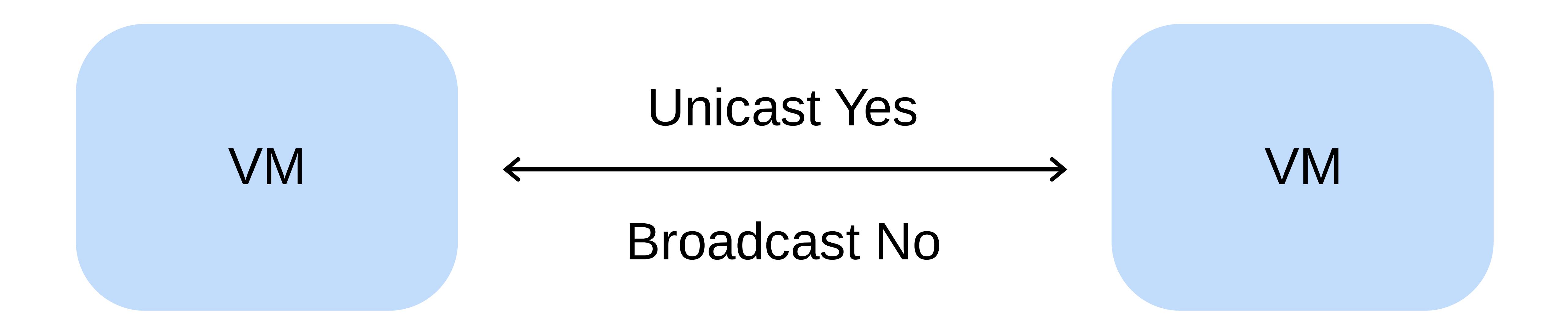
Рекомендации

- Резервируйте compute-ресурсы вашего приложения в разных зонах доступности
- Используйте для продакшена гарантированные выделенные CPU: это даст вам предсказуемость масштабирования. Уплотняйте нагрузки, которым надо меньше 1 CPU, с помощью контейнеризации
- Мониторьте новости о выходе новых платформ, чтобы не оказаться в ситуации, когда кончатся ресурсы текущей.

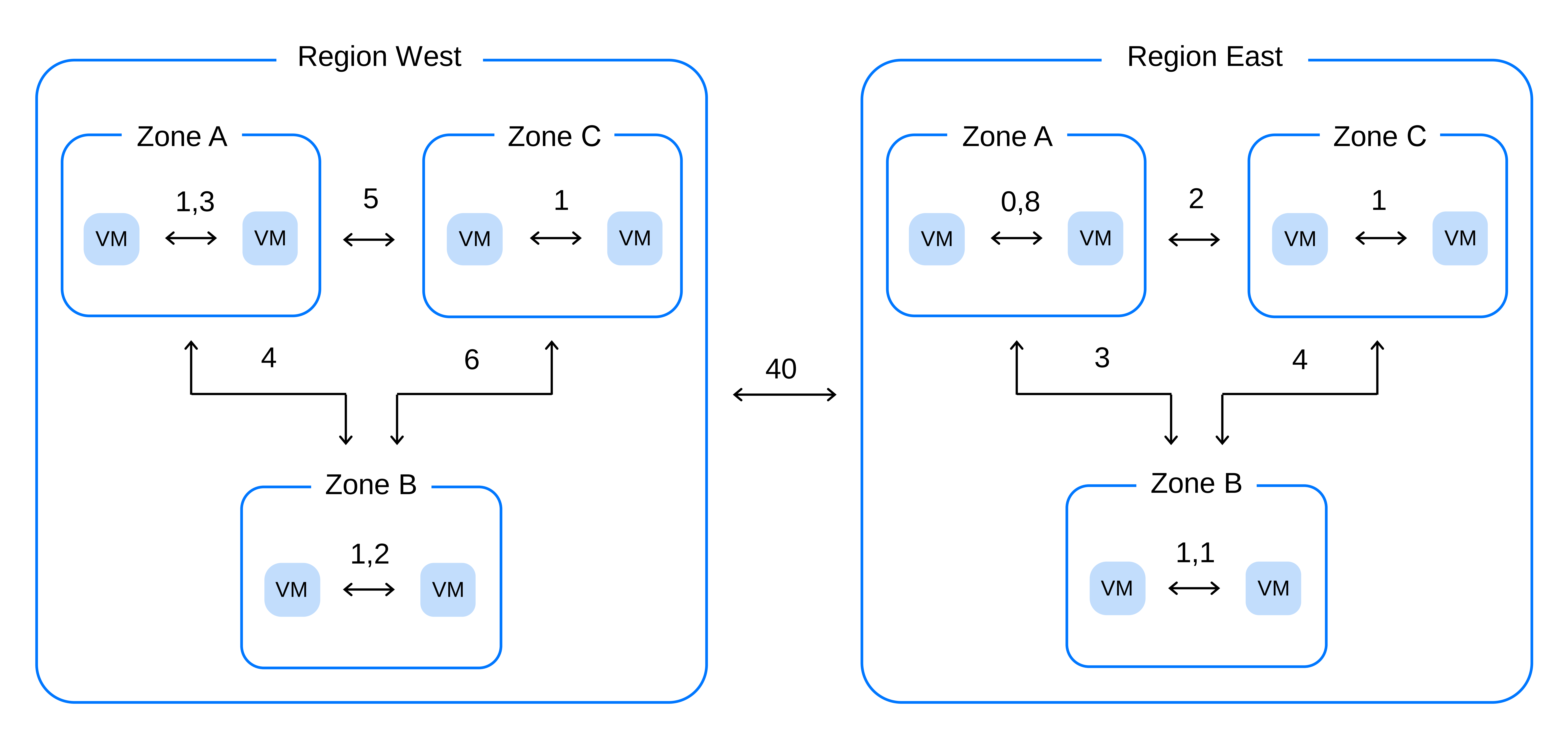


clck.ru/Son8T

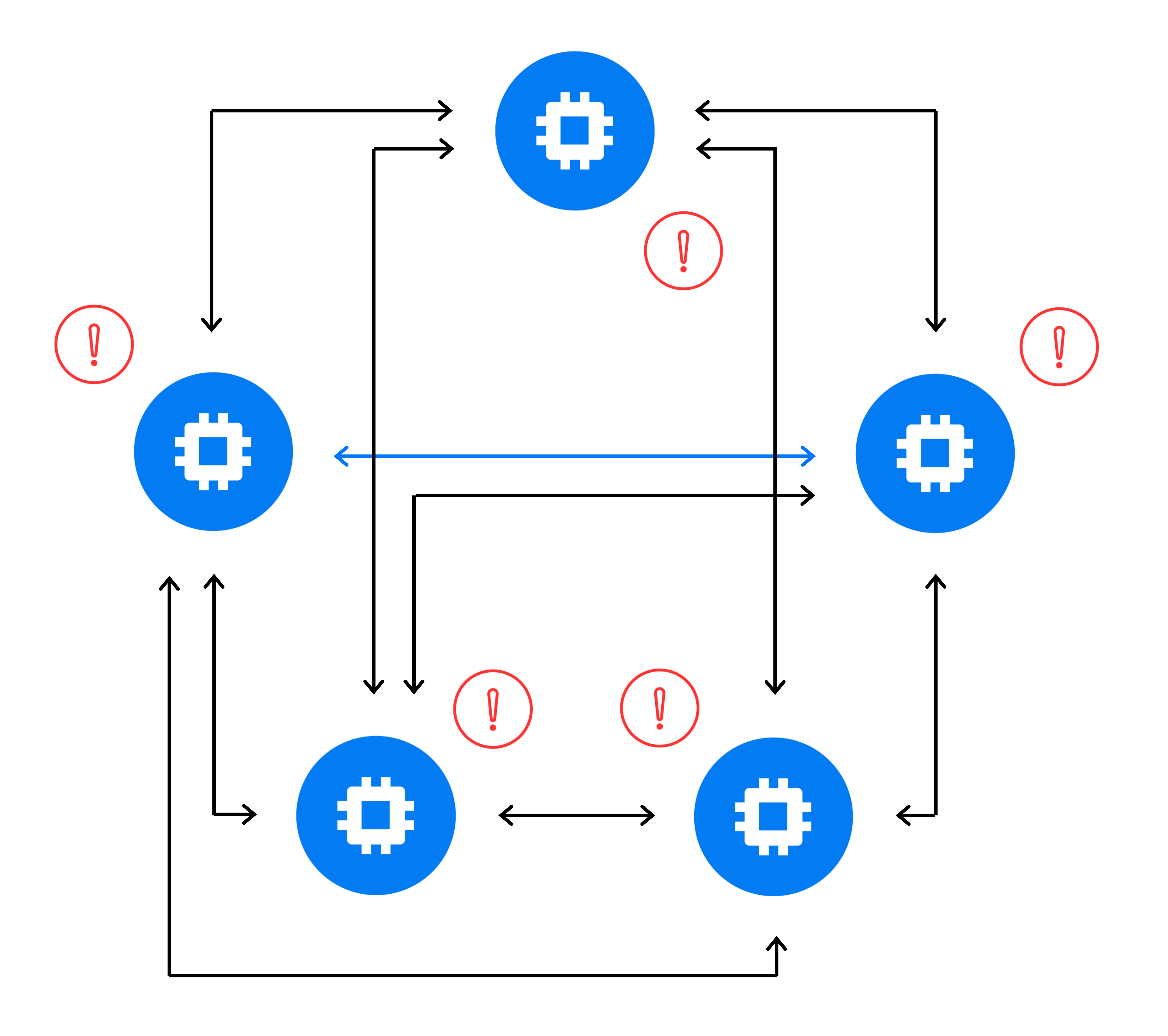
Сеть VPC



Сеть VPC latency



Сеть VPC Tier



Сеть VPC

Особенности

- Уникальная в каждом облаке. Не во всех облаках есть L2 (точнее, есть частично только в AWS)
- Состоит из уровня SDN и физических сетей, настроенных по своим правилам
- Обычно существуют разные tier производительности сетей

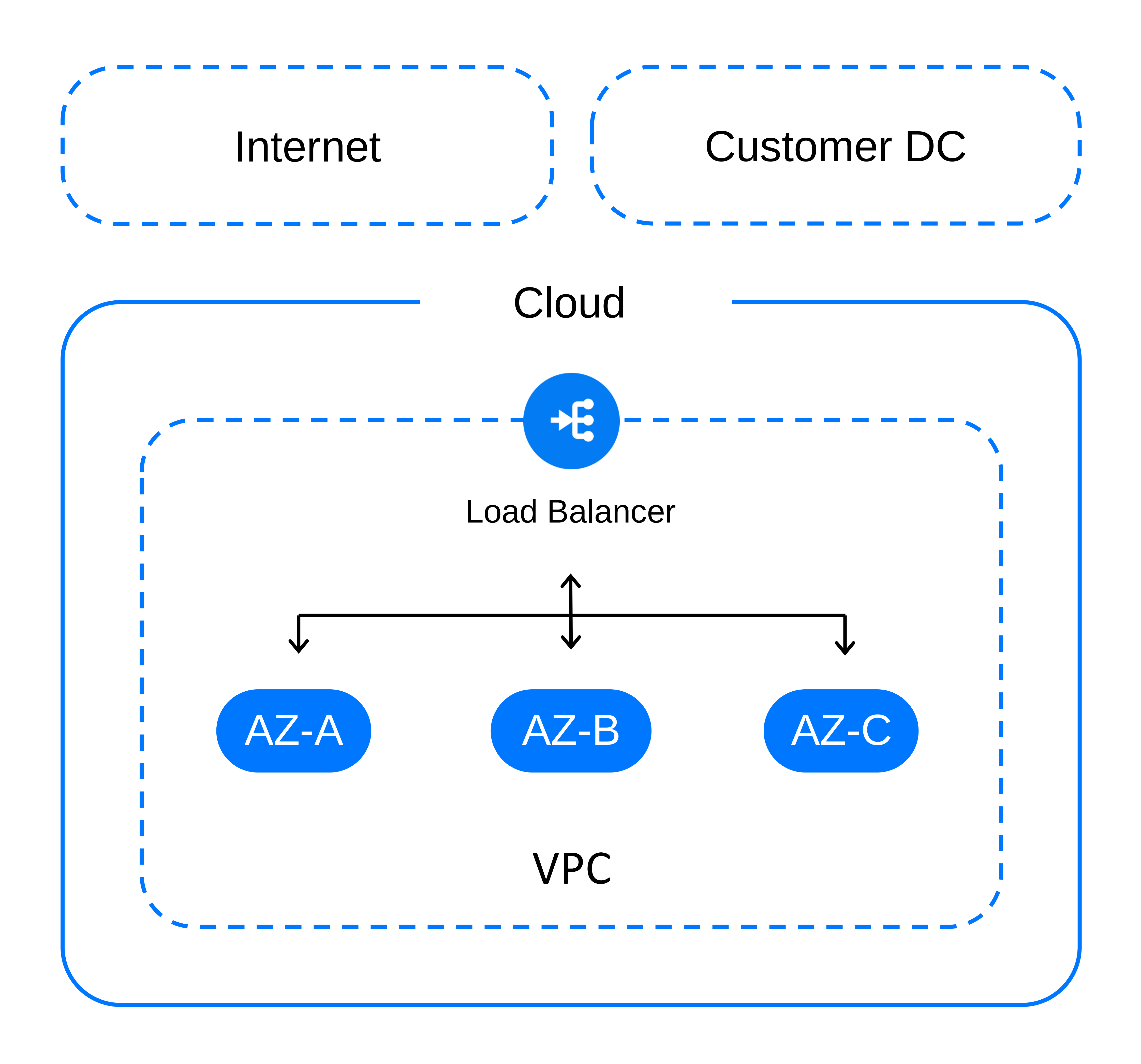
Рекомендации

- Изучите правила работы сети, поддерживаемые протоколы
- Изучите, какие задержки вам стоит ожидать в зонах доступности, между зонами, между регионами. Очертите границу синхронной и асинхронной репликации для ваших систем
- Сделайте нагрузочное тестирование: это позволит понять, требуется ли повышенная производительность сети

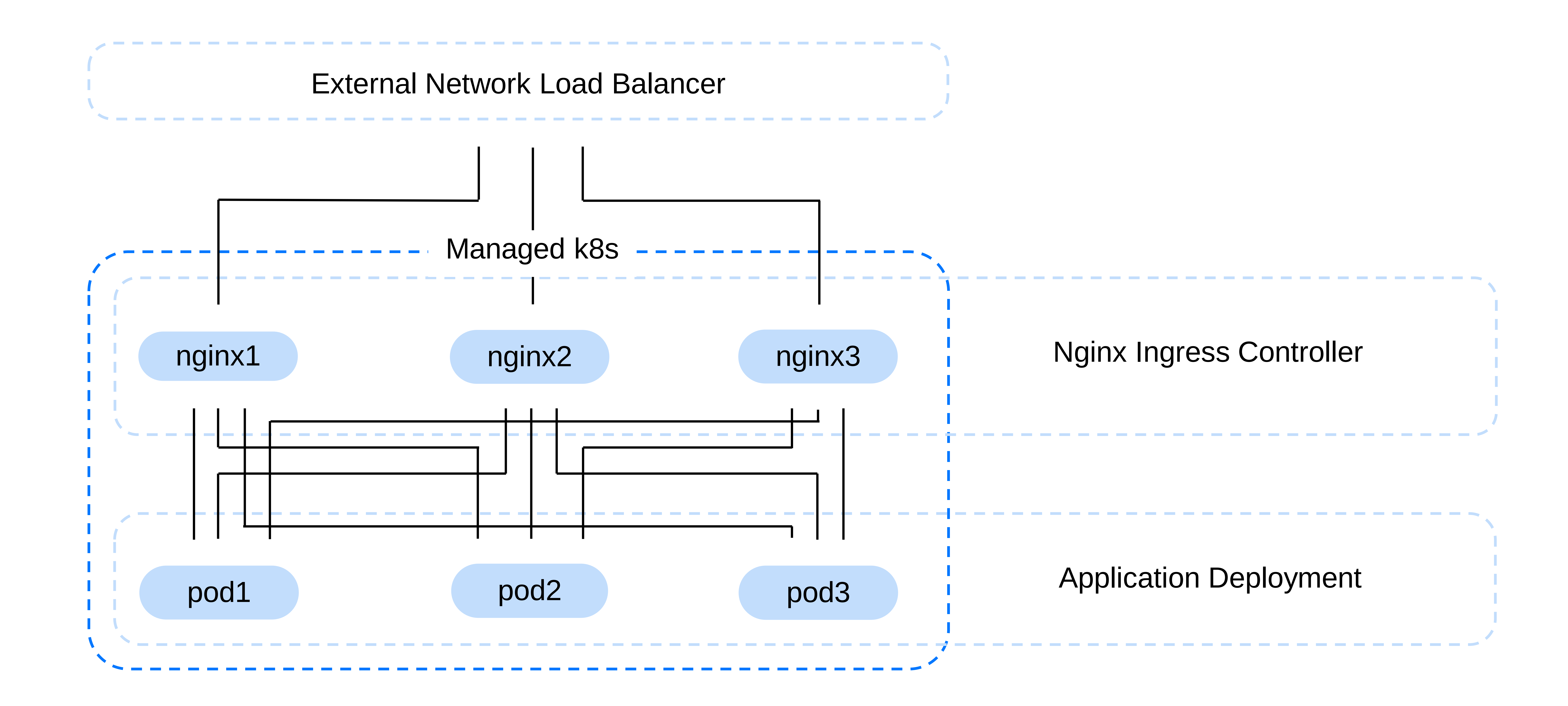


yandex.ru/dev/tank/

Сеть Load Balancer



Сеть Load Balancer



Сеть Load Balancer

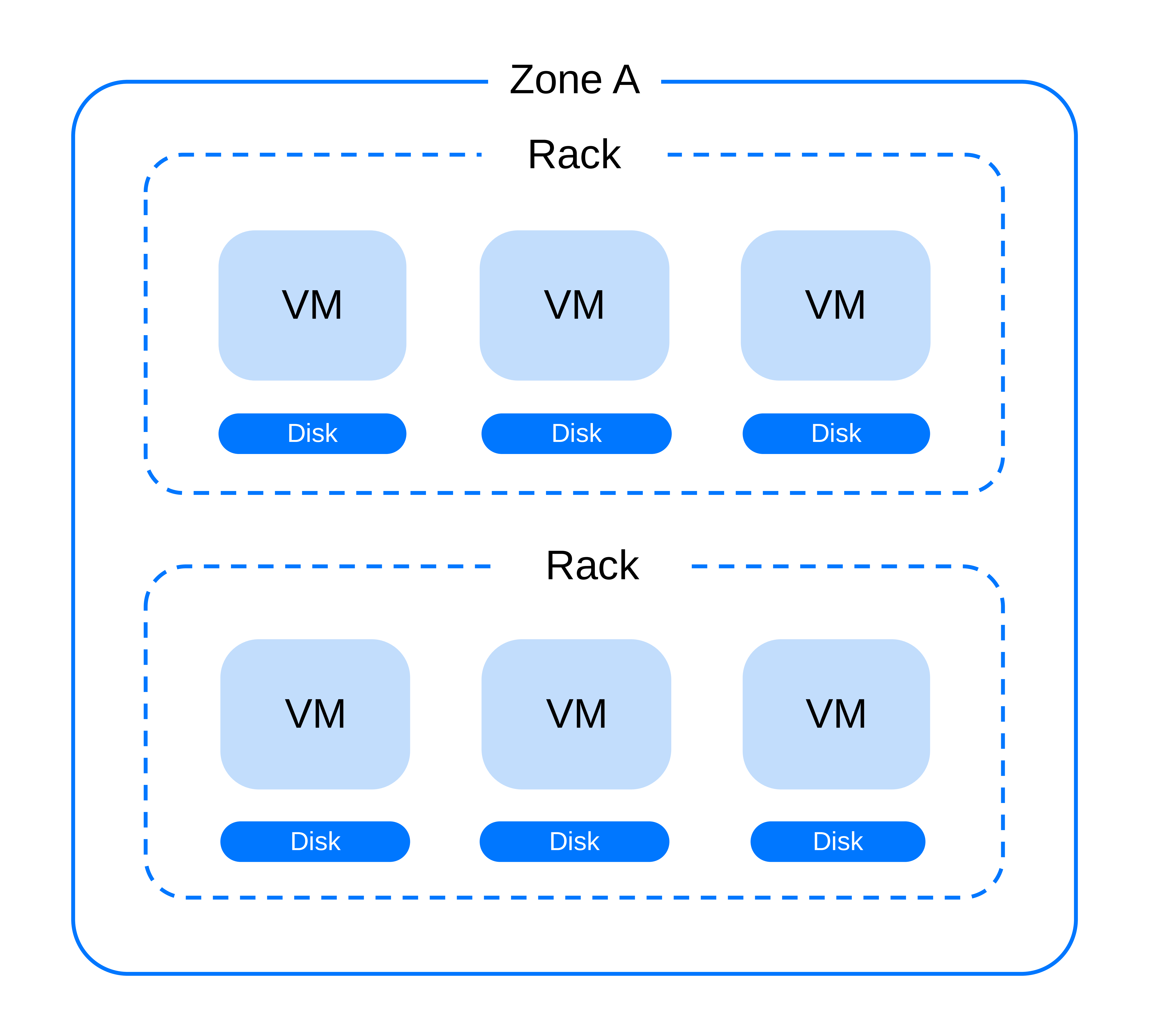
Особенности

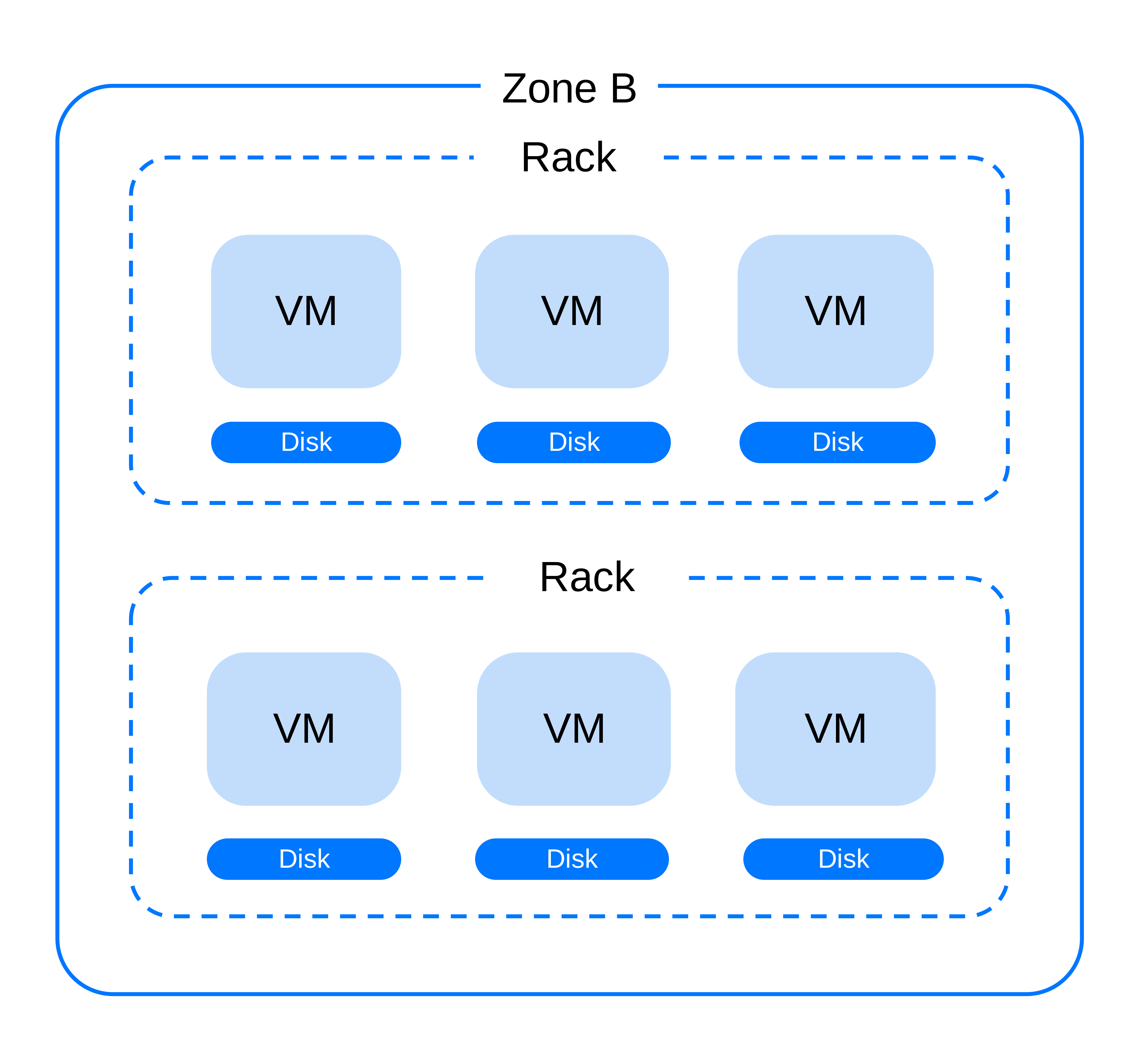
- Уникальный набор в каждом облаке — даже в вопросе scope. Есть внутренние и внешние балансеры
- Обычно бывают:
  - Сетевые балансировщики — работают на уровне L3/L4. «Тупые», но линейно скейлятся и недорогие
  - Application балансировщики — работают на уровне L7. «Умные», но обычно скейлятся лесенкой и могут дорого стоить

Рекомендации

- Изучите типы балансировщиков в облаке, их scope, поддержку протоколов
- По типам балансировщиков
  - Вы не сможете жить без сетевых балансировщиков в облаке
  - Application Load Balancer опционален, тут вам выбирать — брать у облака и крутить свой

Блочное хранилище

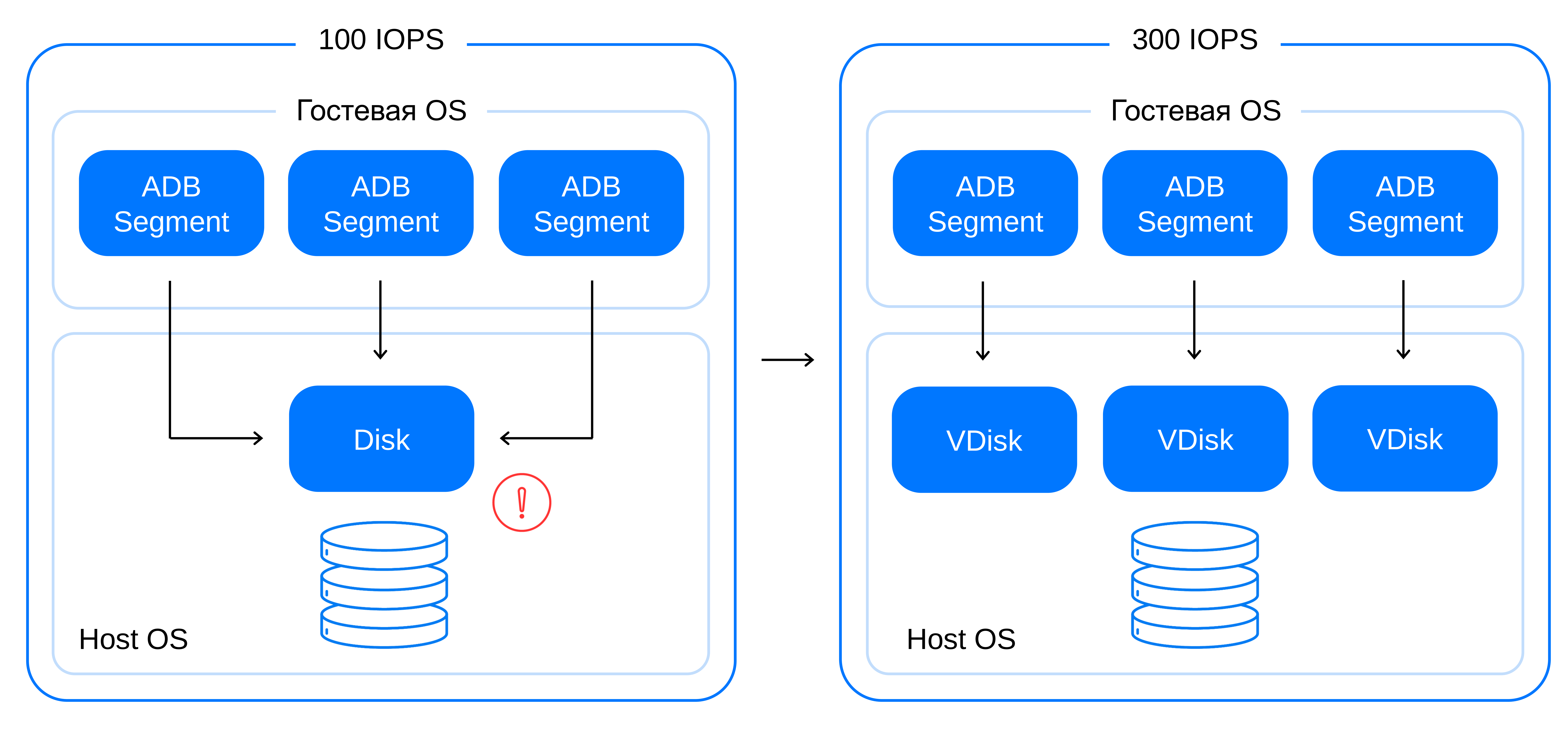




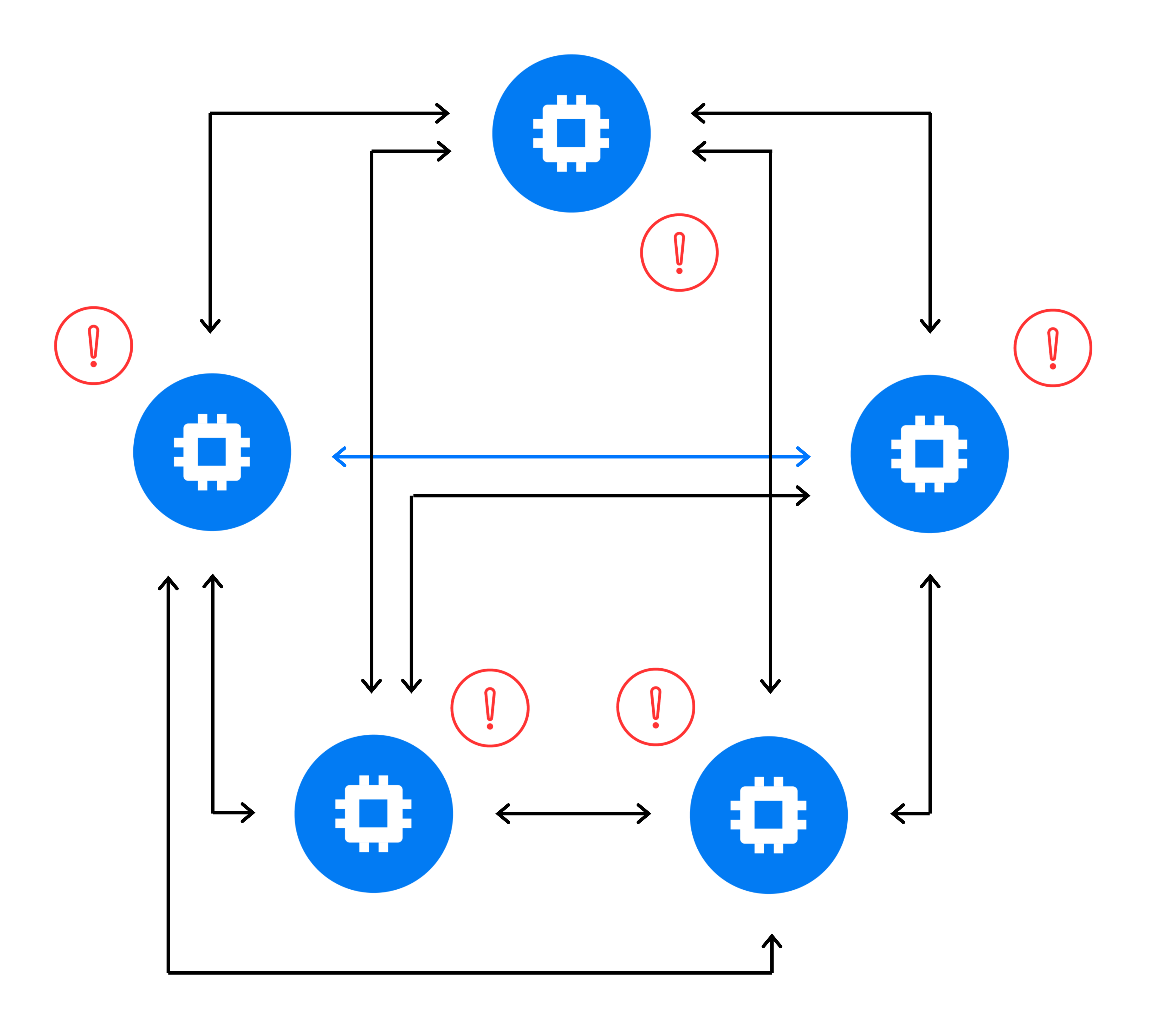
Блочное хранилище



Блочное хранилище — агрегация



Блочное хранилище — Tier



Блочное хранилище

Особенности

- Зональные
- Лимитированы в производительности в зависимости от размера. Лимит (это верхняя граница) — это не гарантия
- Per VM лимиты выше, чем per disk. А иногда их нет
- Есть разные специальные диски для high IO — либо Premium Tier, либо local SSD

Рекомендации

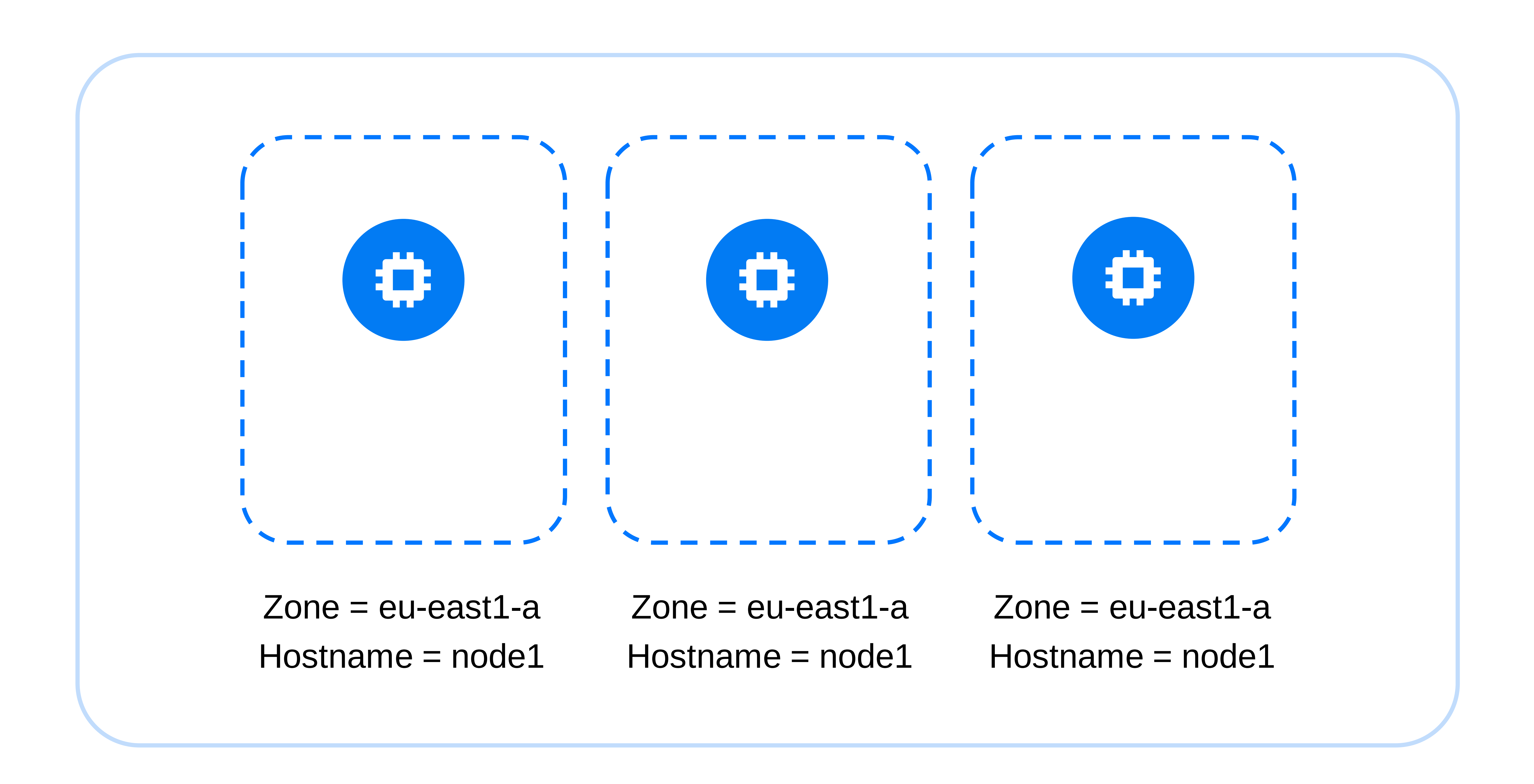
- Реплицируйте statefull-приложения, как минимум в другую AZ
- Не хватает перформанса диска — увеличьте его размер
- Диски можно агрегировать на уровне ОС (RAID 0 или средствами софта)
- Сделайте нагрузочное тестирование: это позволит понять, требуется ли повышенная производительность диска



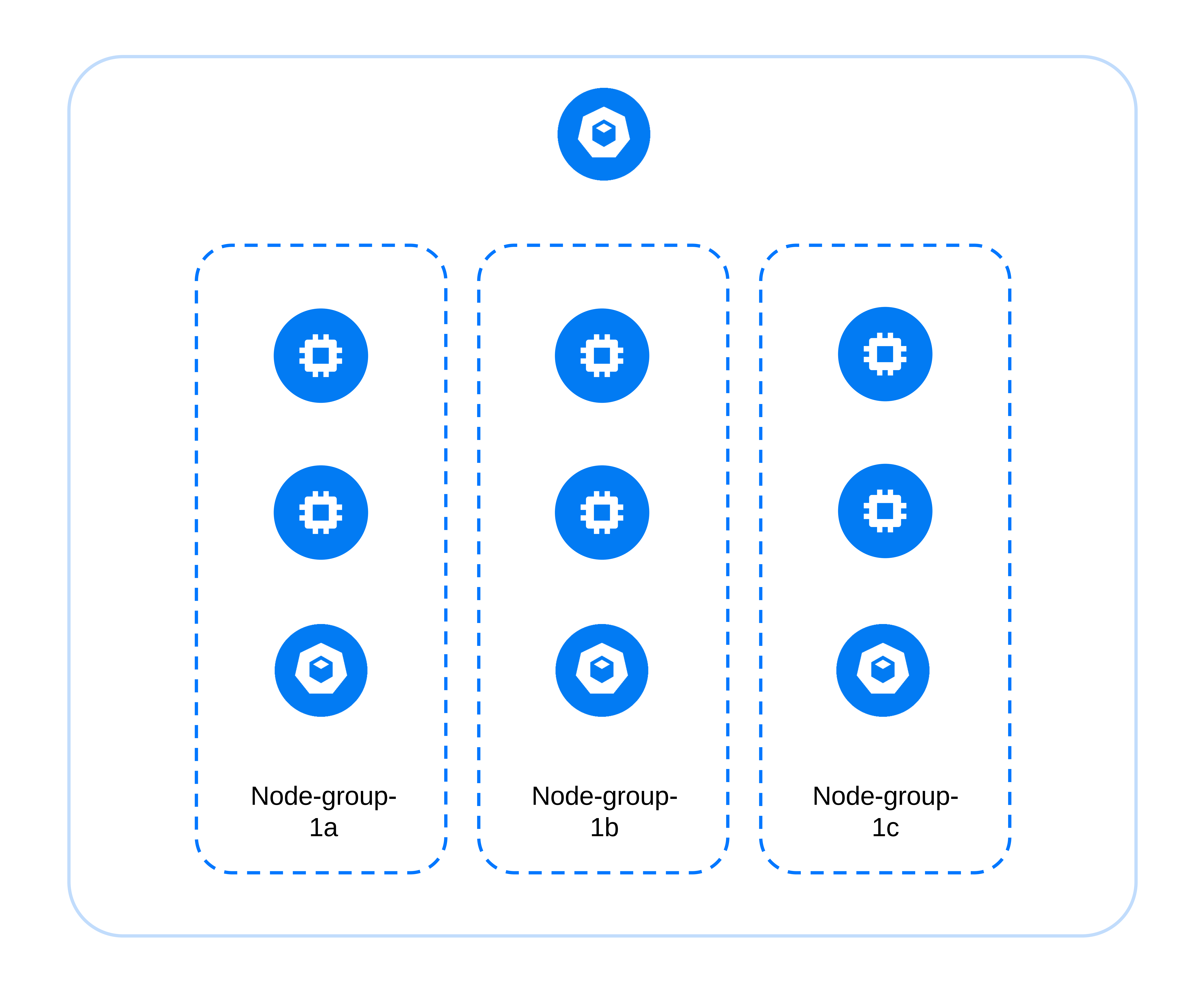
clck.ru/SonGM

Чек-лист про PaaS

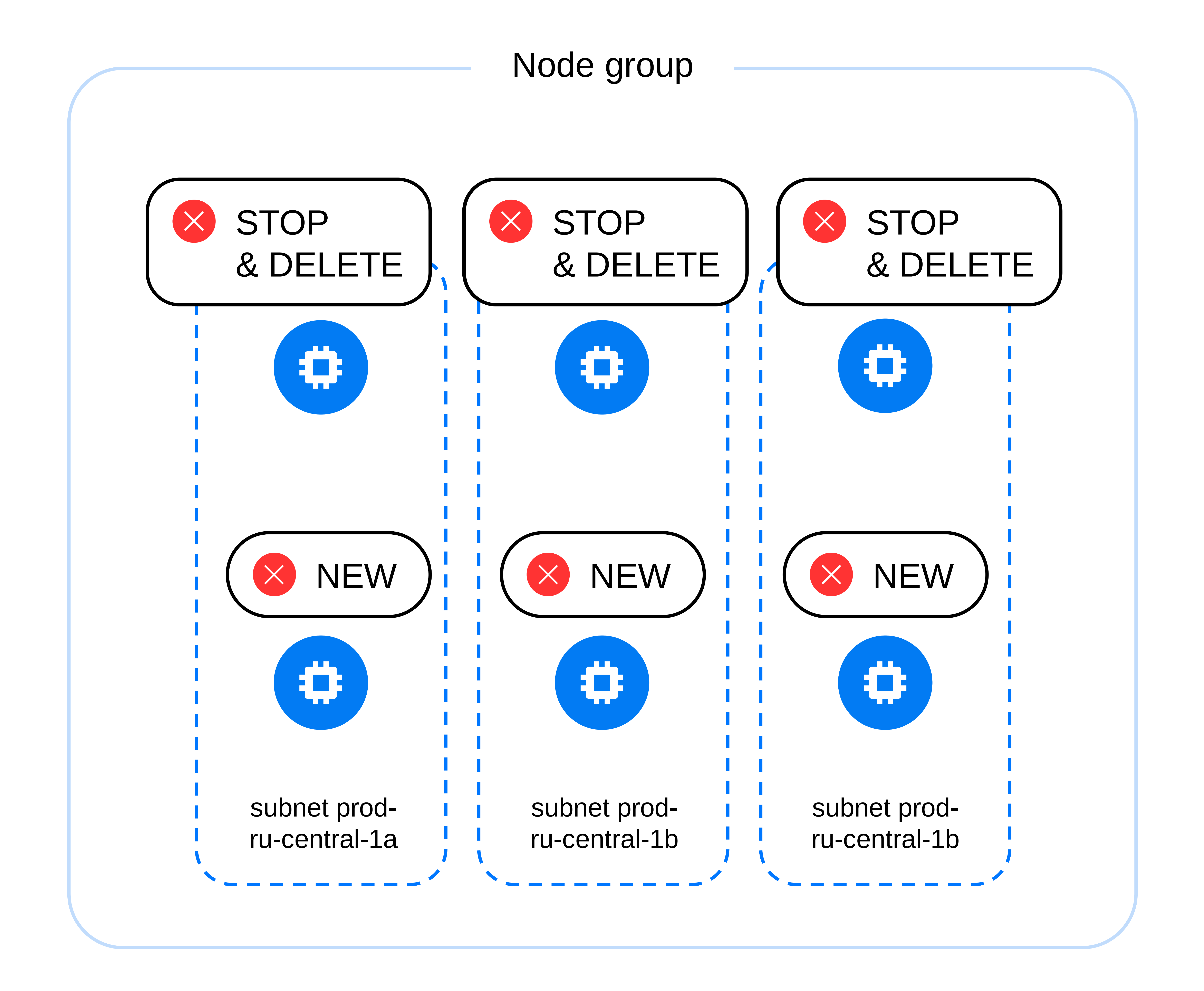
Kubernetes Zone Labes



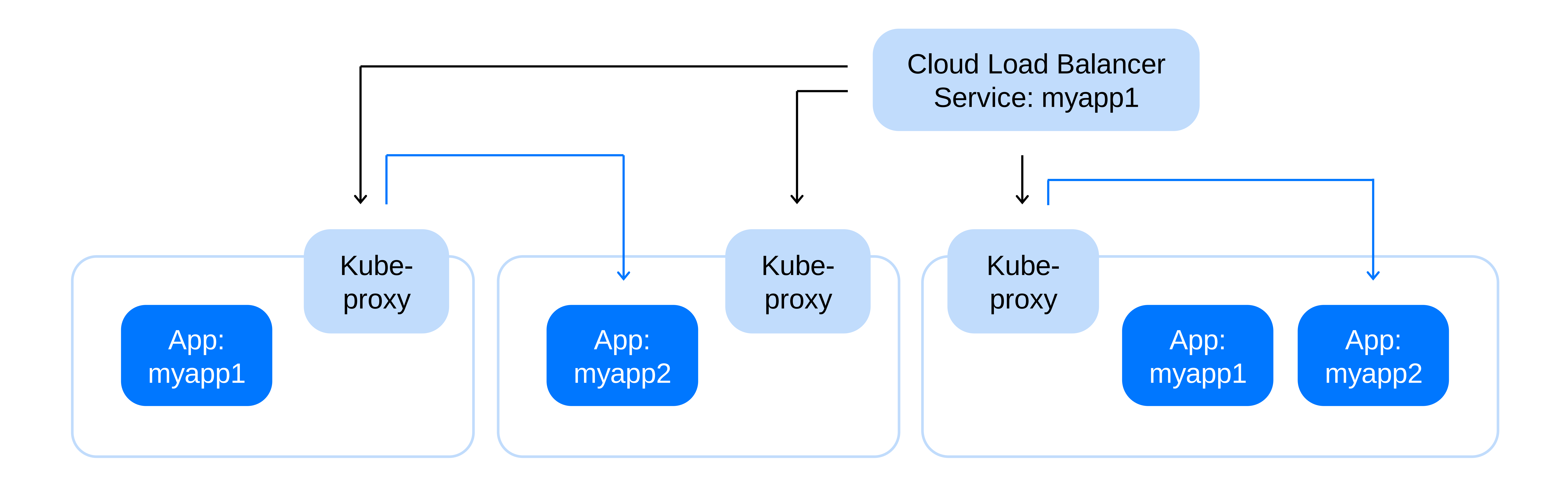
Kubernetes Cluster Autoscaler



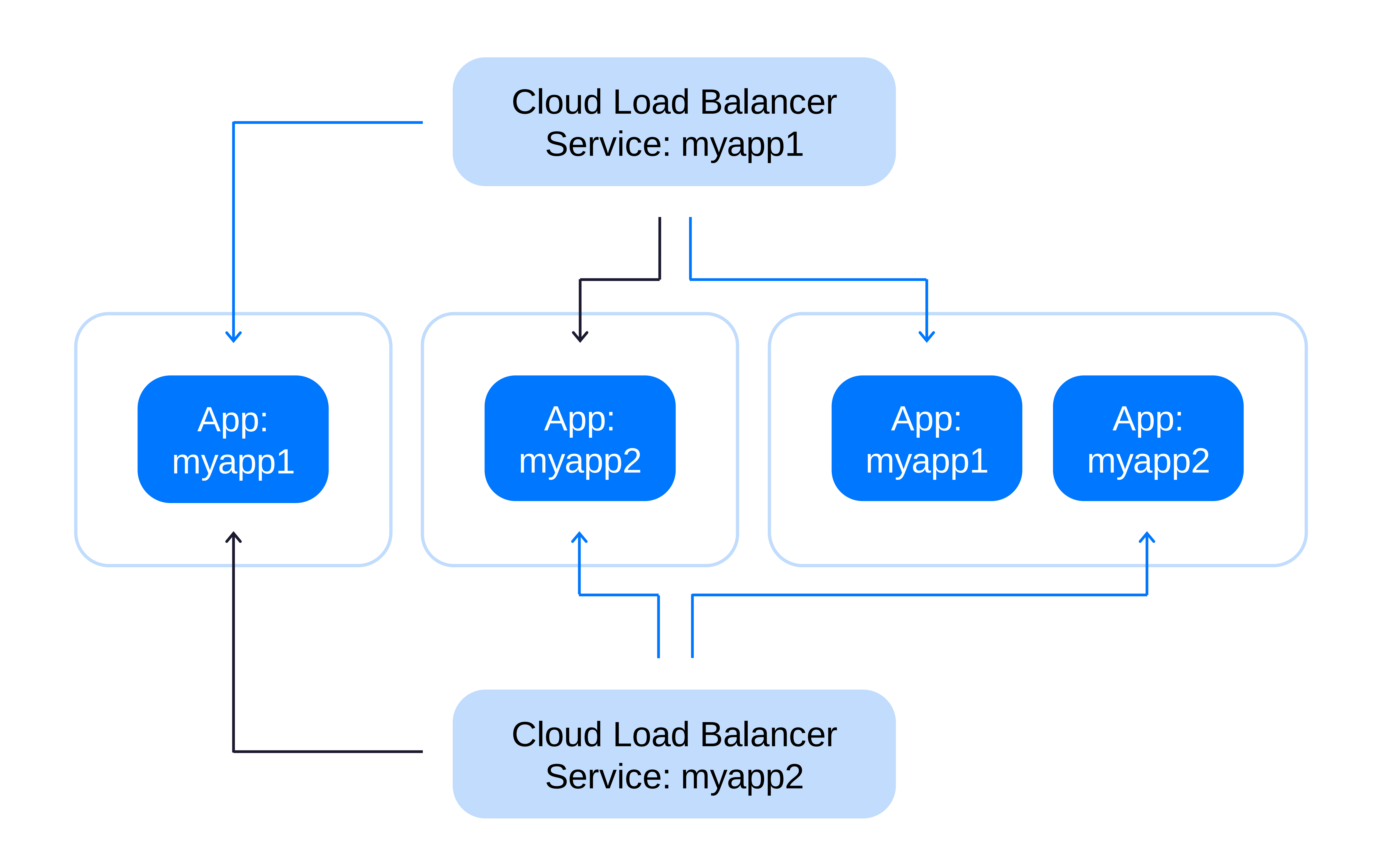
Kubernetes update



Cluster External Traffic Policy



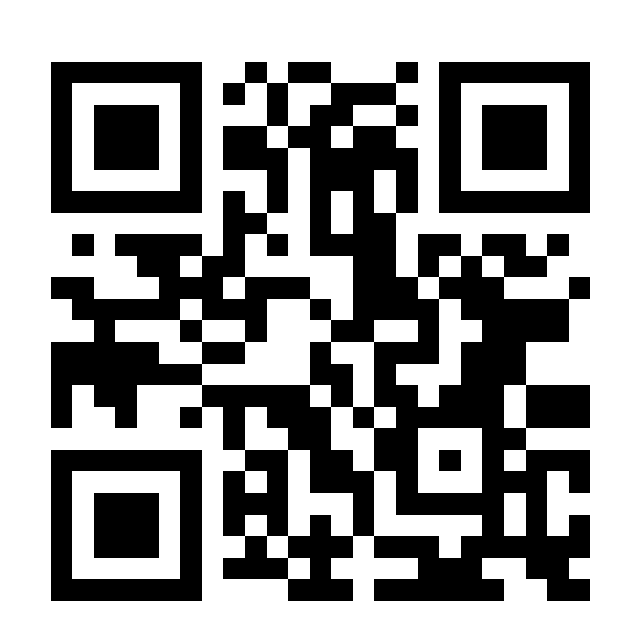
Local External Traffic Policy



Kubernetes

Особенности

- Managed Kubernetes автоматом помечает ноды облачными лейблами — зонами доступности и т. д.
- Cluster Autoscaler тоже умеет оперировать зонами доступности
- Обновление в Managed Kubernetes — это часто создание новых нод, drain и удаление старых нод
- Интеграция с балансировщиком по дефолту создает Cluster-режим балансировки



clck.ru/SonKW

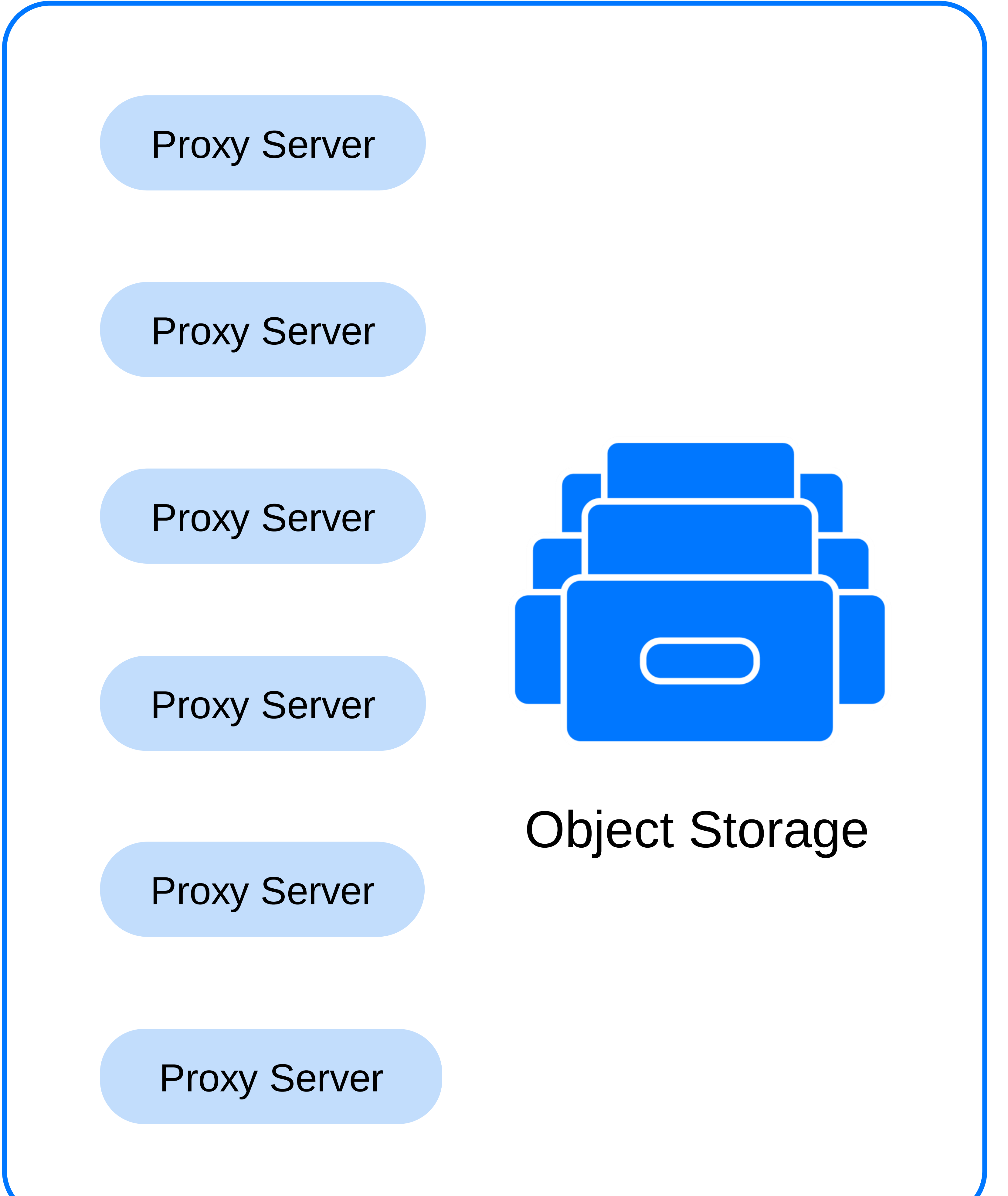
Рекомендации

- Делайте anti-affinity-правила, исходя из ограничений:
  - Можно раскидать deployment’ы без дисков по разным нодам
  - Но StatefulSet,который использует диски, возможно, придётся раскидывать по разным зонам доступности, чтобы сервис работал при выходе из строя одной AZ
- Комбинируйте affinity-правила с Cluster Autoscaler для равномерного скейлинга по зонам доступности
- Используйте podDistruptionBudget и настройки Node Deployment Policy для минимизации даунтаймов при апгрейде узла. Меняйте RollingUpdateStrategy для минимизации даунтаймов при апдейте deployment
- Подключайте Local External Traffic Policy механизм интеграции с балансировщиком, он уменьшит latency ваших пользовательских запросов

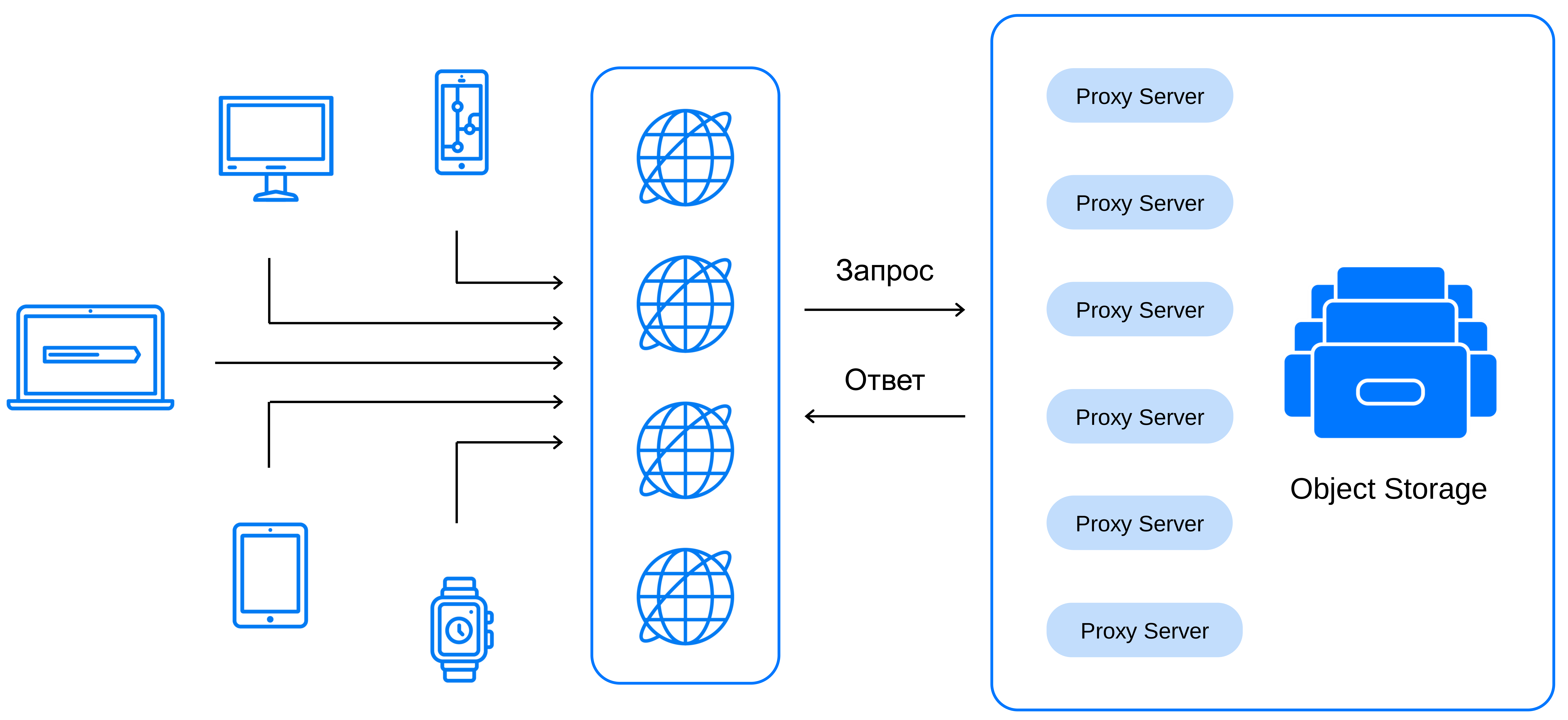
Object Storage

GET/

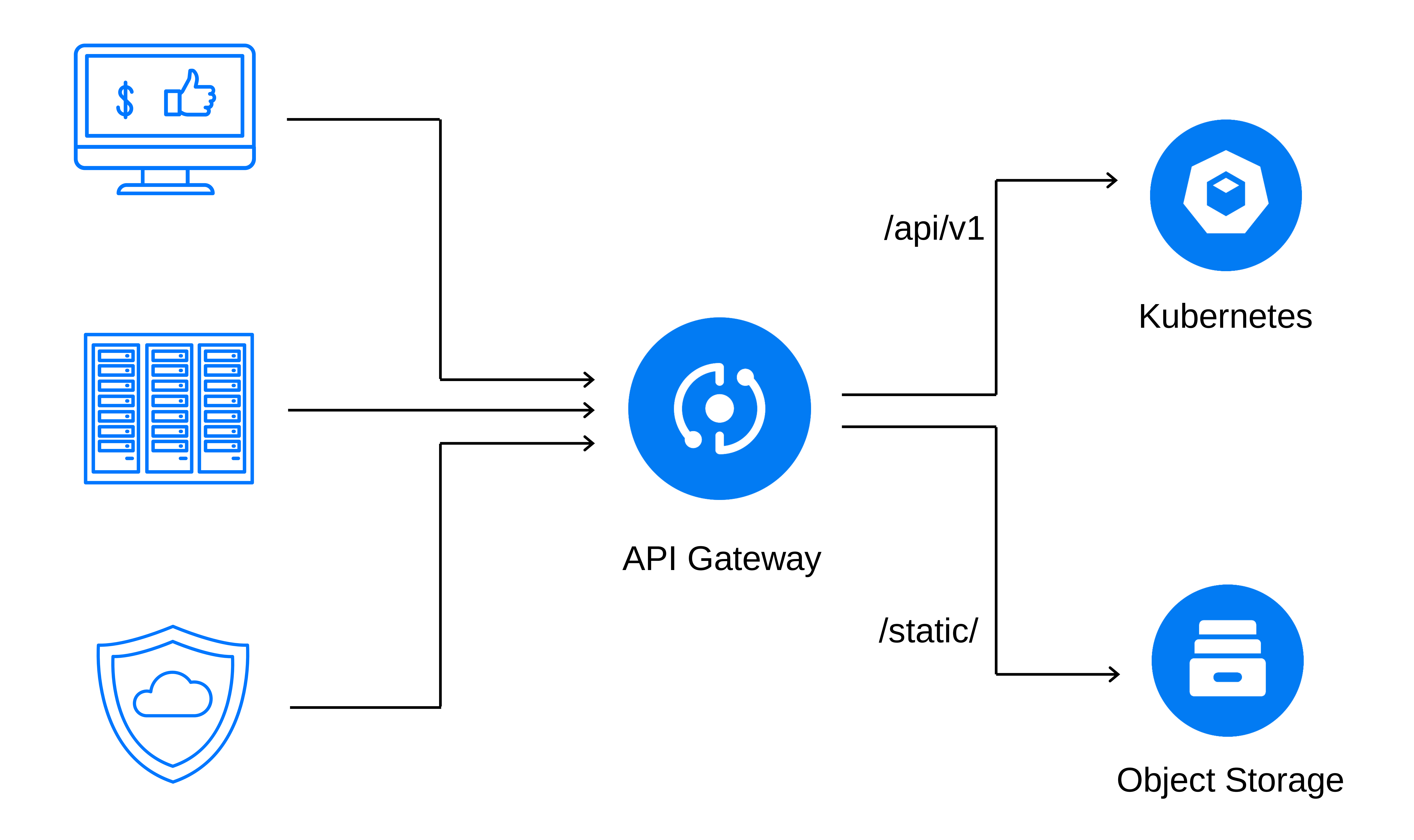
200 OK



Object Storage



Object Storage — офлоад нагрузки



Object Storage

Особенности

- Это не файловое хранилище
- Работает как HTTP/S Endpoint.
  S3-протокол используйте не везде
- Бесконечно масштабируется, но будет брать за это деньги

Рекомендации

- Пользуйтесь SDK для работы с Object Storage, не надо использовать fuse: он by design будет работать медленно
- Офлоадите весь статический контент в storage. В некоторых облаках можно еще добавлять в storage свой домен и TLS-сертификат
- Трафик из Object Storage может быть дорогим — поэтому используйте CDN

Databases













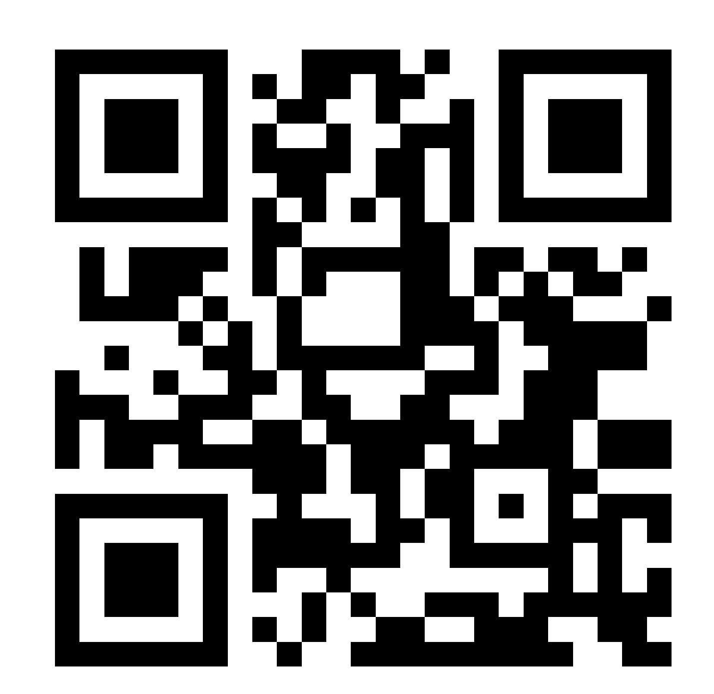
Databases

Особенности

- Есть opensource-варианты баз.
  Они preprovisioned и имеют ряд ограничений.
  Обычно могут горизонтально масштабироваться
- Есть проприетарные базы — они обычно serverless, масштабируются как хотят, но могут быть vendor lock

Рекомендации

- Внимательно изучите ограничения managed базы в рамках
  - Лимитов
  - Возможностей и средств масштабирования Compute и Storage
  - Наличия инструментария автомасштабирования
- Внимательно изучите интерфейсы и лимиты проприетарной СУБД, чтобы принять решение об её использовании или нет



clck.ru/SonMJ



Спасибо!



Нарек Татевосян

Архитектор

nrkk@yandex-team.ru

t.me/nar3k

cloud.yandex.ru